

Nondestructive classification of soft rot disease in napa cabbage using hyperspectral imaging analysis

Hyeyeon Song

World Institute of Kimchi

So-Ra Yoon

World Institute of Kimchi

Yun-Mi Dang

World Institute of Kimchi

Ji-Su Yang

World Institute of Kimchi

In Min Hwang

World Institute of Kimchi

Ji-Hyoung Ha (✉ hajee@wikim.re.kr)

World Institute of Kimchi

Article

Keywords:

Posted Date: May 31st, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1664746/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Identification of soft rot disease in napa cabbage, an essential ingredient of kimchi, is challenging at the industrial scale. Therefore, nondestructive imaging techniques are necessary. Here, we investigated the potential of hyperspectral imaging (HSI) processing in the near-infrared region (900–1,700 nm) for classifying napa cabbage quality using nondestructive measurements. We determined the microbiological and physicochemical qualitative properties of napa cabbage for intercomparison of HSI information, extracted HSI characteristics from hyperspectral images to predict and classify freshness, and established a novel approach for classifying healthy and rotten napa cabbage. Golay's derivative method for data preprocessing was implemented, followed by wavelength selection using variable importance in projection scores. For multivariate data of the classification models, partial least square discriminant analysis, support vector machine (SVM), and random forests were used for predicting cabbage conditions. The SVM model accurately distinguished the cabbage exhibiting soft rot disease symptoms from the healthy cabbage. This study presents the potential of HSI systems for separating soft rot disease-infected napa cabbages from healthy napa cabbages using the SVM model, especially under the most effective wavelengths (970, 980, 1,180, 1,070, 1,120, and 978 nm), prior to processing. These results are applicable to industrial multispectral images.

Introduction

Pectinolytic enterobacteria of the genus *Pectobacterium* (e.g., *Pectobacterium carotovorum* subsp. *carotovorum*, PCC) are the most common cause of major diseases, such as soft rot disease, in healthy vegetables, causing serious deterioration in quality of various healthy produce¹. Soft rot is an important attribute that significantly influences cabbage texture through enzymatic maceration². In addition, sticky patches, a common symptom of spoilage, are the root cause of unpleasant odors and may adversely affect adjacent products. Therefore, research has been performed to identify plant pathogens associated with soft rot disease and to develop protocols for detection and classification of infected vegetables^{3,4}. In addition, PCC has long-lasting biological properties in environmental matrices such as surface water, groundwater, soil, and infected agricultural products such as seeds and vegetables⁵. Recently, Kang et al. reported that PCC contamination has seriously affected quality of kimchi due to the deterioration of commercially cut healthy napa cabbage (*Brassica rapa* L. subsp. *pekinensis* Hanelt)⁶.

Napa cabbage is an essential ingredient of kimchi, a traditional Korean fermented food standardized by the Codex Alimentarius Commission (CXS 223–2001) in 2001⁷. It accounts for 70–80% of kimchi ingredients, thus directly affecting the quality of kimchi⁸. To improve kimchi quality, identifying infected napa cabbage before its processing is crucial. Generally, during the kimchi manufacturing process, soft rot disease of cabbage can be clearly determined while halving the cabbage; however, identification of soft rot disease symptoms in napa cabbage by the naked eye is challenging because of the large amount of cabbage that passes through the cutting machine. Moreover, even if an experienced employee visually inspects the softness of the cabbage tissue, efficiency issues may arise because of the unclear criteria of the operator. Therefore, developing rapid, nondestructive, and unbiased detection methods based on imaging analysis techniques for maintaining kimchi quality is necessary.

Research related to monitoring technologies for preventing soft rot symptoms has been conducted. Numerous studies have reported that several plant pathogens (e.g., *Botrytis cinerea*, *Fusarium sambucinum*, *Pectobacterium carotovorum* ssp. *atrosepticum*, *P. carotovorum* ssp. *carotovorum*, *Phytophthora infestans*, and *Pythium ultimum*) produce volatile markers that can be used to identify fruit and vegetable infections. Morath et al. reported that volatile compound (VC) emanation by fungi has biotechnological potential in the control of postharvest decay⁹. Strobel demonstrated the application of VCs for controlling postharvest fruit diseases using *Muscodor albus*¹⁰. Extensive evidence suggests that the VCs generated by antagonist bacteria could be effective in controlling postharvest decay caused by plant pathogens;

however, those studies often remain unreported^{11,12}. Several studies have demonstrated that plant pathogenic bacteria can alter the profiles of VCs emitted from healthy vegetable tissues^{13,14}. Interestingly, these biological characteristics can be considered specific disease markers. Several monitoring studies on VC emission using nondestructive methods have been conducted; however, few studies have monitored and classified the pathogenic symptoms of PCC-infected postharvest cabbage using imaging analysis techniques based on an active sensing system.

Hyperspectral imaging (HSI) is an innovative platform technique that can integrate spectroscopy¹⁵ and computer vision¹⁶ and simultaneously provide information on the spatial and spectral properties of samples. It has been widely employed in the evaluation of food safety and quality^{17,18}, to study defects or bruise classification^{19,20}, firmness, and soluble solid content²¹, and to monitor pear quality²². As a nondestructive technology, HSI has also been applied for early detection of fruit diseases in apples and peaches^{16,20}. Although several lab-scale studies based on HSI have been performed to classify the spectral characteristics of bean, potato, tomato, and lettuce diseases^{23,24}, no study has focused on soft rot in napa cabbage.

In this study, we investigated the potential of HSI processing techniques at different spectral ranges based on near-infrared (NIR) region for classifying napa cabbage quality using nondestructive measurement. The specific objectives were: determining the physicochemical and microbiological quality properties of napa cabbage for intercomparison of HSI information, screening the HSI characteristic of 900 to 1,700 nm spectral ranges for evaluating napa cabbage freshness, and suggesting a novel approach for the classification of healthy and rotten napa cabbage. Consequently, the spectral preprocessing algorithms were verified, and different classification algorithms, such as partial least square discriminant analysis (PLS-DA), support vector machine (SVM), and random forests (RF), were used for rapid classification of soft rot symptoms from random napa cabbage spectra.

Results And Discussion

Values of microbiological and physicochemical properties

Evaluation of PCC population in napa cabbage samples

After storage, changes in the microbial and physicochemical properties of napa cabbage were analyzed, and the experimental results were used to determine potential soft disease symptoms in the cabbage. The mean titer of PCC in the prepared napa cabbage samples in F2, P1, and P2 sample groups was 1.41 ± 0.62 , 3.61 ± 0.23 , and 5.95 ± 0.27 log colony forming unit/g, respectively (Fig. 2A), whereas viable PCC was not detected in healthy cabbage group (F1). PCC can exist in healthy produce without causing rot^{2,30}. It does not grow in contaminated agricultural products unless they are exposed to certain conditions, such as high temperature or external physical forces (e.g., frictional force and shearing force). Thus, although contaminated with PCC, agricultural products will not show symptoms of soft rot disease. The PCC detected in the healthy cabbage group stored at 30°C has likely proliferated during the storage period but has remained in the latent state. Moreover, the discrepancy in the mean titer of PCC between the P1 and P2 groups is presumed to be due to the storage temperature. It has been reported that the optimal growth temperature for *Pectobacterium* spp. is within the range from 20°C to 34°C^{3,31}. The differential PCC mean values in the four groups of napa cabbage are considered suitable in a reasonable range for separating healthy samples, infected samples, infected but healthy samples, and samples that are healthy but contain PCC bacterial cells.

Evaluation of water content in napa cabbage samples

Figure 2B presents the descriptive statistics of the water content measured in the four groups of napa cabbage. The highest water content was detected in P2, and it was significantly different from that in other groups; the water content

was relatively uniform in both healthy cabbage groups (F1 and F2) and infected cabbage groups (P1 and P2), ranging from 85.4–93.8%. The difference in water content of napa cabbage has been attributed to the mean values of the PCC titer—the higher the mean values of the PCC titer, the higher the water content. The variation in water content in various napa cabbage samples could be related to the maceration of cabbage tissues^{4,32}.

Evaluation of VOCs in napa cabbage samples

For specific VOC detection as a marker of soft rot disease, cabbage samples stored for three days at designated temperatures were exposed to the headspace, including the SPME fiber. Table 1S shows the various VOCs emitted by the four types of napa cabbage sample groups detected by HS-SPME-GC-MS analysis. Among cabbage samples with confirmed presence of PCC (F2, P1, and P2), only the P2 group emitted 2,3-butanediol as a volatile marker (Fig. 3). Using HS-SPME combined with GC-MS analysis, Yang et al. demonstrated that 2,3-butanediol is a specific volatile metabolic marker of soft rot disease symptoms in PCC-infected cabbage. The 3-hydroxy-2-butanone pathway is required for *P. carotovorum* pathogenesis, and 2,3-butanediol plays an important role in the production of volatile products¹. Furthermore, annotation using the KEGG database revealed that butanoate metabolism, which contributes to the rotting scent, is associated with metabolic pathways in *P. carotovorum* subsp. *carotovorum* strain PCC21 (isolated from *B. rapa* ssp. *pekinensis*) (Fig. S1). In the present study, to clearly differentiate the sample group with soft rot disease symptoms, all groups were analyzed using VOC fingerprinting by headspace capillary-gas chromatography-ion mobility spectrometry (HS-GC-IMS). HS-GC-IMS assay for VOC profiling is highly efficient and an automatable analytical technique with acceptable sensitivity to VOCs³³. Based on the representative HS-GC-IMS spectral information of the napa cabbage samples (Fig. S2A1-6 and B), 11 target VOC fingerprinting spots were determined to be reliable markers for differentiating soft rot disease symptoms. These VOC fingerprinting spots were investigated in the unsaturated region of the IMS spectra to ensure the validity of the experimental data. Multivariate data were analyzed using supervised principal component analysis (PCA), which was performed considering Cattell's scree test, Kaiser's eigenvalue-one criterion, and Bartlett's test of sphericity (SC). According to Tabachnick et al., Kaiser–Meyer–Olkin (KMO) test values above 0.5 are considered acceptable, above 0.9 are excellent, above 0.8 are great, and above 0.7 are good³⁴. In this study, the KMO value was above 0.9, considering the interdependence of the components in the PCA. The PCA score plot was constructed to distinguish between the soft rot symptom group (P2) and the non-soft rot symptom group (F1, F2, and P1) based on the 11 selected VOC fingerprinting spots, which were differentiated with a separation of 98.1 (Fig. S2C).

Spectral images analysis

Changes in the microbial and physicochemical properties of napa cabbage were analyzed to determine whether the cabbage had soft disease symptoms. A total of 197 bands between 938 and 1,711 nm were selected as effective bands owing to obvious noise in the rear and front regions of the spectral data. Figure 4A shows a raw spectrum plot and Fig. 4B a mean raw spectrum plot from the selected ROI. Similar spectrum patterns were observed in all napa cabbage samples without any noticeable spectral differences. However, the spectral reflectance of the P2 group was lower than those of the F1, F2, and P1 groups (Fig. 4B) when compared with the mean spectral plot. The napa cabbage samples had absorption peaks at approximately 1,180 nm and 1,450 nm, which could have been induced by plant stress triggered by various external factors (e.g., bacterial infection, light, or temperature). Particularly, phytopathogenic bacteria can be responsible for changes in relative water content and cellulose³⁵. According to Andro et al., PCC generates various enzymatic substances such as cellulase, extracellular pectinases, and proteases that degrade numerous plant cell wall components³⁶. Moreover, cellulase and extracellular pectinases are major virulence factors for the development of soft rot symptoms, which then promote hydration of plant cell tissues³⁷. Therefore, it was concluded that the plant disease mechanism of PCC is correlated with the spectral characteristics of napa cabbage samples.

Classification models based on selected wavelengths

A PLS-DA classification model was developed to distinguish between the soft rot symptom group and the non-soft rot symptom group of napa cabbage. For this purpose, it is important to prepare a cabbage sample in which each characteristic is clearly defined. Four groups of napa cabbage samples were prepared. According to the microbiological and physicochemical properties, F1, F2, and P1 samples were designated as the non-soft rot disease symptom groups, and P2 samples were designated as the soft rot disease symptom group. Although data preprocessing is not preferred for developing models based on raw data for industrial applications, data preprocessing methods were used to ensure the reliability of the models in this study. The classification of the soft rot and non-soft rot symptom groups of napa cabbage was based on S. Golay's 2nd derivative. The data were also pre-processed using Golay's derivative method (Fig. S3). The results of the preprocessing were less noisy than the results obtained after single processing using Golay's 2nd derivative and those obtained using the mean normalization. Only a few optimal wavelengths that convey the most important information representing the entire spectrum were selected to reduce the high dimensionality of the extracted spectral data. The VIP method was used to reduce the number of wavelengths in the model classifier and to increase the accuracy of the real-time classification model. We presented the most important wavelengths using the PLS-DA model that efficiently classify soft rot disease infection, as shown in Fig. 5. Wavelengths in the HSI range were 970, 980, 1,180, 1,070, 1,120, and 978 nm, and they are associated with water absorption, water sensitivity, total chlorophyll, texture, and internal chemical composition^{38,39}. Particularly, texture, cellulose, and water sensitivity are key factors for evaluating the soft rot symptom^{36,40}. The point just above the VIP score '1' in the SWIR plot was neglected because it fell in the poor SNR region. The number of variables was reduced from 197 to 6 wavelengths in the HSI range on the basis of variable selection for HSI using the VIP scores (Fig. 5). Table 1 presents the confusion matrices of the model with selected wavelengths and their ability to discriminate soft rot samples from non-soft rot samples for the three algorithms (SVM, PLS-DA, and RF) as classifiers. Confusion matrices in the three classifiers indicate that the classification of soft rot napa cabbages was acceptable. The SVM model showed 99% SENS, 96% SEPC, and 99% ACC for calibration sets and 96% SENS, 88% SEPC, and 95% ACC for validation sets of the napa cabbages (Table 1). The overall cross-validation analysis confirmed that the SENS, SEPC, and ACC were significantly superior in the SVM models than in PLS-DA and RF.

Table 1

Confusion matrices for discrimination of napa cabbages inoculated with *Pectobacterium carotovorum* subsp. *carotovorum* (PCC) and healthy napa cabbages based on the pre-processed data in hyperspectral wavelength ranges.

SVM	Confusion matrix (%)			SENS ¹⁾	SPEC ²⁾	ACC ³⁾	CV ⁴⁾	
	Calibration	PV ⁶⁾	AV ⁵⁾					ACC
			0	1			96%	
			0	132	2	99%	96%	99%
			1	0	45			
	Validation	PV	AV					
			0	1				
			0	43	2	96%	88%	95%
			1	1	14			
PLS-DA	Confusion matrix (%)			SENS	SPEC	ACC	CV	
	Calibration	PV	AV					ACC
			0	1			92%	
			0	129	6	96%	88%	96%
			1	1	44			
	Validation	PV	AV					
			0	1				
			0	41	4	91%	78%	92%
			1	1	14			
Random forest	Confusion matrix (%)			SENS	SPEC	ACC	CV	
	Calibration	PV	AV					ACC
			0	1			89%	
			0	129	6	96%	88%	95%
			1	3	42			
	Validation	PV	AV					
			0	1				
			0	39	6	87%	68%	87%
			1	2	13			
¹⁾ SENS: Sensitivity, ²⁾ SPEC: specificity, ³⁾ ACC: accuracy, ⁴⁾ CV: calibration value, ⁵⁾ AV: actual values, and ⁶⁾ PV: predicted values								

The PLS method is the most commonly employed chemometric approach for multivariate data analysis. Furthermore, PLS in classification models is a practical and powerful method based on binary classification of highly correlated variables⁴¹. The PLS-DA model has been previously evaluated in similar studies and has been recognized and

recommended for its high accuracy and simplicity^{42,43}. RF has recently received considerable attention and has been considered an ensemble classification approach with demonstrated superiority and accuracy⁴⁴. However, a major drawback of RF is that it does not work for datasets with insufficient features⁴⁵. Recently, the SVM model was recognized for its extremely high accuracy compared with that of PLS-DA⁴⁶. Its reputation has been steadily increasing in the biochemical field and has been considered an optimal classifier in many studies because of its extremely high accuracy⁴⁷. The SVM model is advantageous compared with other models because of its capacity to treat high-dimensional and nonlinear data. In particular, an SVM model based on the structural risk minimum (SRM) provides an easy-to-compute solution, is flexible, and shows high generalization capacity⁴⁸. Zhang et al. reported that SVM models built on the reflectance of HSI in the short-wave infrared range (874–1,734 nm) could classify the ripeness of strawberries (unripe, mid-ripe, and ripe) at the optimal wavelengths obtained from the loadings of PCA models⁴⁹. Huang et al. demonstrated that the SVM model is a more accurate classifier than the PLS-DA model for classifying mealiness in apples (mealy vs. non-mealy apples)⁵⁰. Similarly, in the present study, the model generated by SVM was superior to PLS-DA and RF in the classification of soft rot-infected napa cabbages.

Conclusions

Herein, we report chemometric tools combined with an active sensing system based on HSI spectra (874–1,734 nm) for rapid detection and efficient classification of soft rot disease symptoms in napa cabbage. The proposed method successfully classified even PCC-infected cabbage samples that did not emit 2,3-butanediol as a marker volatile substance of soft rot disease. Particularly, although there were no specific symptoms that could be observed with the naked eye, the SVM model accurately distinguished the soft rot disease symptom group. Consequently, the experimental results of this study suggest the potential of HSI using the SVM model, especially under the most effective wavelengths (970, 980, 1,180, 1,070, 1,120, and 978 nm), in separating napa cabbage infected with soft rot from healthy napa cabbages prior to processing. In addition, it was demonstrated that these results can be successfully applied to multispectral images for industrial purposes.

Materials And Methods

Napa cabbage preparation

Strains of PCC KACC 18645 isolated from napa cabbage was kindly donated by the Rural Development Administration (Deokjin-gu, Jeonju-si, Korea). Healthy napa cabbage was purchased from a local agricultural wholesale market (Gwangju, Korea). The napa cabbages were divided into four groups (30 per group; total of 120 whole cabbages) for conducting the following four treatments: healthy napa cabbage stored at 5°C (F1), healthy napa cabbage stored at 30°C (F2), napa cabbage inoculated with PCC and stored at 5°C (P1), and napa cabbage inoculated with PCC and stored at 30°C (P2). High temperature (30°C) was used to simulate the summer ambient conditions. The storage period lasted 3 days. The weight of each napa cabbage was $4,000 \pm 200$ g. For artificial PCC inoculation (P1 and P2), the cut side of the bottom of napa cabbage (core part) was immersed in PCC suspension (approximately $5 \log_{10}$ colony forming unit/mL) for 2 h at 20°C. The healthy napa cabbages (F1 and F2) were immersed in sterile water for 2 h at 20°C. Subsequently, each cabbage was cut into two equal sections (total of 60 pieces per treatment) (Fig. 1). In each treatment group (F1, F2, P1, and P2), 45 pieces were used for calibration groups (total of 180 samples; F1: 45 samples, F2: 45 samples, P1: 45 samples, and P2: 45 samples) to develop the classification models and 15 pieces for the validation group (total of 60 samples; F1: 15 samples, F2: 15 samples, P1: 15 samples, and P2: 15 samples) to verify the fitness of the calibration models. For diagnostic analyses of the soft rot disease in cabbage, viable PCC cells were detected using PMA/RT-qPCR assay²⁵, water content was determined using the gravimetric method²⁶, and extracellular metabolites of the soft rot

microorganisms were identified using headspace solid-phase microextraction followed by gas chromatography coupled with mass spectrometry (HS-SPME-GC-MS)⁴.

Analysis of microbiological and physicochemical properties

Enumeration of PCC populations and water content

Each cabbage section was mixed with 30 mL of PBS in a sterile filter stomacher bag (Seward Limited, London, UK) and evenly blended using a stomacher (SH11 M; Elmex, Tokyo, Japan) for 60 s. For bacterial genomic DNA extraction, 20 mL of eluate containing PCC bacterial cells was collected and centrifuged at 8,000 rpm for 15 min to concentrate the planktonic cells. A Power-Prep quick DNA extraction kit (Kogene-Biotech Co., Seoul, Korea) was used to purify the extracted bacterial genomic DNA according to the manufacturer's instructions. The suspension samples of the purified DNA were stored at -70°C until further use, and 5 μL of each test sample was used directly for one-step real-time reverse transcription PCR (RT-qPCR). For RT-qPCR amplification, a PowerChek *Pectobacterium carotovorum* Real-Time PCR kit (Kogene-Biotech Co.) was used to amplify a DNA template in a 7500 Fast Real-Time PCR System (Applied Biosystems, Foster City, CA, USA), equipped with Applied Biosystems 7500 Fast Real-Time PCR detection system software (version 3.0; Applied Biosystems), according to the manufacturer's instructions. The standard curve was prepared by illustrating the mean threshold cycle (C_t) value ($n = 3$) obtained using the serial dilutions of a standard PCC DNA transcript for RT-qPCR to quantify the PCC bacterial cells in the cabbage sample; a linear plot was obtained, with a coefficient of determination (R^2) > 0.9988 and a slope of -3.239 , reflecting adequate correlation in the tested range between 1.57 and 6.58 \log_{10} genomic DNA copies. The water content of each cabbage sample was measured using an infrared moisture analyzer (Model MB 45; Ohaus, Pine Brook, NJ, USA).

VOC analysis of napa cabbage samples

The VOCs released from the samples by the headspace solid-phase microextraction (HS-SPME) fiber were analyzed using a gas chromatography system (Agilent 7890A; Agilent Technologies, Santa Clara, CA, USA) coupled with a mass spectrometer (5977 B; Agilent Technologies). Headspace solid-phase microextraction (HS-SPME) was performed using a multipurpose autosampler (MPS2; Gerstel, Mülheim an der Ruhr, Germany). VOC fingerprinting analysis was performed using gas chromatography-ion mobility spectrometry (FlavourSpec®; G.A.S., Dortmund, Germany).

Hyperspectral imaging data acquisition

Hyperspectral imaging system

A line-scanning type (push-broom scanner) NIR-HSI system (N17E-QE; SPECIM, Spectral Imaging Ltd., Oulu, Finland) was employed to collect the HSI data for napa cabbage samples. The NIR-HSI system used for the analysis was composed of an SWIR camera equipped with an OLES56 camera lens (SWIR-CL-400-N25E; SPECIM) covering the spectral wavelength range of 874–1734 nm with 320×256 pixels, a spectral resolution of approximately 12 nm, and a translating scanner. The system consisted of a plate conveyor driven by a stepper motor (Isuzu Optics Corp, Zhubei, Taiwan) and two 150 W tungsten halogen lamps (Fi ber-Lite DC950 Illuminator; Dolan Jenner Industries Inc., Boxborough, MA, USA) fixed symmetrically on both sides of the camera at a 45° angle as the illumination source. The system was placed in a dark room and was controlled using a computer. To scan each sample, each spectral image was acquired for 5 min under controlled ambient temperature (20°C) in the dark chamber.

Image acquisition and correction

The exposure time of the camera, scanning speed of the plate conveyor, and distance between the napa cabbage sample and lens are the main factors influencing the HSI data. Therefore, to collect non-deformable and clear images containing whole napa cabbage samples, three main factors must be accurately set. In this study, the main influencing factors were

controlled using LUMO® software (SPECIM). The scanning speed and distance between the lens and samples were set as 22.5 mm/s and 31 cm, respectively. For image correction, both the hyperspectral image acquisition and white/dark reference images were acquired under the same experimental conditions. To remove dark noise and uneven illumination, both hyperspectral image acquisition and white/dark reference images were acquired under the same experimental conditions. A piece of white Teflon (99% reflectance) was used to acquire the white reference image, and the dark reference image was captured by turning off the light source and completely covering the camera lens using an opaque cap. Finally, the corrected images were calculated using the following equation:

$$I_C = \left[\frac{I_O - I_D}{I_W - I_D} \right] \times 100$$

where I_C is the calibrated image, I_O is the original hyperspectral image, I_D is the dark reference image, and I_W is the white reference image. Sample spectral images were extracted using ENVI 4.7 software (Research Systems Inc., Boulder, CO, USA). For hyperspectral images ranging from 938 to 1710, the region of interest (ROI) must be predefined for the extraction of spectral information. In this study, the ROI of each sample was determined to be an orthogonal section of each napa cabbage, from which the spectral data were extracted. Finally, the mean spectral information of each sample was acquired by averaging all pixels of the ROI for additional analysis.

Model development and performance evaluation

Data preprocessing and data analysis

Preprocessing was conducted to improve the accuracy of the classification model and eliminate the influence of irregularities in the spectral data induced by sample texture, light scattering, random noise. In this study, Golay's derivative method for data preprocessing was used to ensure the reliability of the models.

Feature selection

PLS-DA preprocessing methods for feature selection were applied to pre-process the spectral data. Selecting the variable with the highest weight is essential for a quick, simple, and efficient implementation of the PLS-DA model in an image-information-based classification system. Important wavelength selection was obtained by estimating variable importance in projection (VIP) scores of PLS-DA models²⁷. Generally, VIP scores < 0.8 represent variables that are less effective, while VIP scores > 1.0 represent highly dominant variables²⁷. All chemometric techniques were implemented using the data processing server MetaboAnalyst 4.0²⁸.

Model development and performance evaluation

Classification models can induce convergence instability because the HSI, as high-dimensional imaging data, presents a high degree of interband correlation and results in data redundancy. To overcome these technical limitations and improve classification stability and implementation of these models in a multispectral imaging system, only optimal wavebands were selected. In this study, the algorithm for the most common way to solve integer programs, such as the branch-and-bound technique, was used to select the optimal wavebands suitable for implementation in a multispectral imaging system. The combination of the preprocessing method and the developed model that evaluated the highest accuracy was accepted to be the most suitable for differentiating between napa cabbage with or without soft rot symptoms. The regression coefficient vectors multiplied by the original masked HSI images were employed with the preprocessing method and model combination to develop the chemical imaging data. The obtained chemical images were converted into binary images using a threshold value. Resulting chemical images based on pixels denoted "0 value" were classified

as non-soft rot symptom group, and those with "1 value" were classified as soft rot symptom group. Binary chemical images were employed to eliminate tiny pixels that caused data misclassification, and the binary chemical images were further clarified. Preprocessing was conducted to increase the accuracy of the classification model and eliminate the bias due to irregularities in the spectral data induced by sample texture, light scattering, and random noise. In this study, prior to the application of classification models, Golay's derivative method was used as a preliminary operation to exclude outliers. Three multivariate models, SVM with linear kernel, PLS-DA, and RF, were developed to classify the soft rot and non-soft rot symptom groups of napa cabbage. Multivariable analysis procedures have been recognized as standard approaches and are widely used in the analysis of HIS images. Hyperspectral data analysis aims to develop a classification or a predictive model. In this study, pre-processed hyperspectral data from individual methods was employed for developing a classification model between the soft rot and the non-soft rot symptom groups of napa cabbage.

Multivariate data analysis

PLS-DA was used as the classifier. According to Barker and Rayens, the PLS-DA model utilizes the interaction between sample characteristics and spectral intensities by maximizing the covariance between variables²⁹. The performance of the developed PLS-DA model was investigated based on sensitivity (SENS), specificity (SPEC), and accuracy (ACC), representing the true positive rates, true negative rates, and overall correctly classified samples, respectively.

$$Specificity = \frac{Truenegatives}{(Truenegatives + Falsepositives)}$$

$$Sensitivity = \frac{Truepositives}{(Truepositives + Falsenegatives)}$$

$$Accuracy = \frac{Truepositives + Truenegatives}{(Truepositives + Truenegatives + Falsepositives + Falsenegatives)}$$

where n is the number of samples. The accuracy of the PLS-DA model can be improved by choosing the optimal number of latent variables (LVs) on the basis of the minimum root mean square error of cross-validation (RMSECV).

Declarations

Acknowledgments

This research was supported by grants from the World Institute of Kimchi (KE2102-2 and KE2202-2) and funded by the Ministry of Science and ICT, the Republic of Korea.

Author contributions

HS and **SRY** performed most of the experiments, contributed to the experimental design, analyzed the data, performed most of the experiments, and wrote the first draft of the manuscript. **HS** and **SRY** contributed equally to this work and are considered co-first authors. **YMD**, **JSY**, and **IMH** performed experiments. **JHH** was responsible for the experimental design, data coordination, analysis and interpretation, writing, revision, and finalization of the manuscript. All the authors have read and approved the final manuscript.

Competing interests

The authors declare no conflict of interest.

Research involving plants

Our study complies with relevant institutional, national, and international guidelines and legislation.

References

1. Marquez-Villavicencio, M. P., Weber, B., Witherell, R. A., Willis, D. K. & Charkowski, A. O. The 3-hydroxy-2-butanone pathway is required for *Pectobacterium carotovorum* pathogenesis. *PLOS ONE* **6**, e22974 (2011). <https://doi.org/10.1371/journal.pone.0022974>.
2. Perombelon, M. C. M. Potato diseases caused by soft rot erwinias: an overview of pathogenesis. *Plant Pathol.* **51**, 1–12 (2002). [https://doi.org/10.1046/j.0032-0862.2001.Short title.doc.x](https://doi.org/10.1046/j.0032-0862.2001.Short%20title.doc.x).
3. Tsrer (Lahkim) et al. Characterization of *Pectobacterium brasiliense* strains from potato and vegetables in Israel. *Plant Pathol.* **70**, 2179–2187 (2021). <https://doi.org/10.1111/ppa.13454>.
4. Yang, J.S., Lee, H.W., Song, H. & Ha, J.-H. Volatile metabolic markers for monitoring *Pectobacterium carotovorum* subsp. *carotovorum* using headspace solid-phase microextraction coupled with gas chromatography-mass spectrometry. *J. Microbiol. Biotechnol.* **31**, 70–78 (2021). <https://doi.org/10.4014/jmb.2009.09028>.
5. Hadas, R., Kritzman, G., Gefen, T. & Manulis, S. Detection, quantification and characterization of *Erwinia carotovora* ssp. *carotovora* contaminating pepper seeds. *Plant Pathol.* **50**, 117–123 (2001). <https://doi.org/10.1046/j.1365-3059.2001.00540.x>.
6. Kang, M. et al. Determination of transfer patterns of *Pectobacterium carotovorum* subsp. *carotovorum* planktonic cells and biofilms during mechanical cutting of kimchi cabbage. *J. Food Sci.* **84**, 2603–2609 (2019). <https://doi.org/10.1111/1750-3841.14749>.
7. Codex Alimentarius Commission. Codex standard for kimchi (CXS 223–2001). Food and Agriculture Organization of the United Nations/ World Health Organization. Preprint at https://www.fao.org/fao-who-codexalimentarius/sh-proxy/en/?Ink=1&url=https%253A%252F%252Fworkspace.fao.org%252Fsites%252Fcodex%252FStandards%252FCXS%2B223-2001%252FCXS_223e.pdf (2001).
8. Jung, M. Y. et al. Role of jeotgal, a Korean traditional fermented fish sauce, in microbial dynamics and metabolite profiles during kimchi fermentation. *Food Chem.* **265**, 135–143 (2018). <https://doi.org/10.1016/j.foodchem.2018.05.093>.
9. Morath, S. U., Hung, R. & Bennett, J. W. Fungal volatile organic compounds: a review with emphasis on their biotechnological potential. *Fungal Biol. Rev.* **26**, 73–83 (2012). <https://doi.org/10.1016/j.fbr.2012.07.001>.
10. Strobel, G. Muscodor species-endophytes with biological promise. *Phytochem. Rev.* **10**(2), 165–172 (2011). <https://doi.org/10.1007/s11101-010-9163-3>.
11. Li, Q. et al. Effects of volatile substances of *Streptomyces globisporus* JK-1 on control of *Botrytis cinerea* on tomato fruit. *Biol. Control* **61**, 113–120 (2012). <https://doi.org/10.1016/j.biocontrol.2011.10.014>.
12. Zheng, M., Shi, J., Shi, J., Wang, Q. & Li, Y. (2013). Antimicrobial effects of volatiles produced by two antagonistic *Bacillus* strains on the anthracnose pathogen in postharvest mangos. *Biol. Control* **65**, 200–206. <https://doi.org/10.1016/j.biocontrol.2013.02.004>.
13. Jansen, R. M. C. et al. Detection of diseased plants by analysis of volatile organic compound emission. *Ann. Rev. Phytopathol.* **49**, 157–174 (2011). <https://doi.org/10.1146/annurev-phyto-072910-095227>.

14. Laothawornkitkul, J. et al. Volatile organic compounds as a diagnostic marker of late blight infected potato plants: A pilot study. *Crop Prot.* **29**, 872–878 (2010). <https://doi.org/10.1016/j.cropro.2010.03.003>.
15. Wang, J.-X. et al. Determination of the moisture content of fresh meat using visible and near-infrared spatially resolved reflectance spectroscopy. *Biosyst. Eng.* **162**, 40–56 (2017). <https://doi.org/10.1016/j.biosystemseng.2017.07.004>.
16. Zhang, B. et al. Computer vision detection of defective apples using automatic lightness correction and weighted RVM classifier. *J. Food Eng.* **146**, 143–151 (2015). <https://doi.org/10.1016/j.jfoodeng.2014.08.024>.
17. Liu, Q. et al. Quantitative visualization of fungal contamination in peach fruit using hyperspectral imaging. *Food Anal. Methods* **13**, 1262–1270 (2020). <https://doi.org/10.1007/s12161-020-01747-x>.
18. Zhang, B. et al. Influence of physical and biological variability and solution methods in fruit and vegetable quality nondestructive inspection by using imaging and near-infrared spectroscopy techniques: a review. *Crit. Rev. Food Sci. Nutr.* **58**, 2099–2118 (2018). <https://doi.org/10.1080/10408398.2017.1300789>.
19. Feng, L. et al. Detection of subtle bruises on winter jujube using hyperspectral imaging with pixel-wise deep learning method. *IEEE Access* **7**, 64494–64505 (2019). <https://doi.org/10.1109/ACCESS.2019.2917267>.
20. Li, J., Chen, L. & Huang, W. Detection of early bruises on peaches (*Amygdalus persica* L.) using hyperspectral imaging coupled with improved watershed segmentation algorithm. *Postharvest Biol. Technol.* **135**, 104–113 (2018). <https://doi.org/10.1016/j.postharvbio.2017.09.007>.
21. Sun, J., Künnemeyer, R., Mcglone, A. & Rowe, P. Multispectral scattering imaging and NIR interactance for apple firmness predictions. *Postharvest Biol. Technol.* **119**, 58–68 (2016). <https://doi.org/10.1016/j.postharvbio.2016.04.019>.
22. Yu, X., Lu, H. & Wu, D. Development of deep learning method for predicting firmness and soluble solid content of postharvest Korla fragrant pear using Vis/NIR hyperspectral reflectance imaging. *Postharvest Biol. Technol.* **141**, 39–49 (2018). <https://doi.org/10.1016/j.postharvbio.2018.02.013>.
23. Cho, H. et al. Hyperspectral determination of fluorescence wavebands for multispectral imaging detection of multiple animal fecal species contaminations on romaine lettuce. *Food Bioproc. Technol.* **11**, 774–784 (2018). <https://doi.org/10.1007/s11947-017-2032-y>.
24. Dutta, M. K., Singh, A. & Ghosal, S. A computer vision based technique for identification of acrylamide in potato chips. *Comput. Electron. Agric.* **119**, 40–50 (2015). <https://doi.org/10.1016/j.compag.2015.10.007>.
25. Jeong, S., Lee, J. Y., Yoon, S., Moon, E. W. & Ha, J. A quantitative PCR based method using propidium monoazide for specific and sensitive detection of *Pectobacterium carotovorum* ssp. *carotovorum* in kimchi cabbage (*Brassica rapa* L. subsp. *pekinensis*). *Lebensm. Wiss. Technol.* **113**, 108327 (2019). <https://doi.org/10.1016/j.lwt.2019.108327>.
26. Nakajima, S., Genkawa, T., Miyamoto, A. & Ikehata, A. Useful tissues in cabbage head for freshness evaluation with visible and near infrared spectroscopy. *Food Chem.* **339**, 128058 (2021). <https://doi.org/10.1016/j.foodchem.2020.128058>.
27. Farrésa, M., Platikanov, S., Tsakovski, S. & Tauler, R. (2015). Comparison of the variable importance in projection (VIP) and of the selectivity ratio (SR) methods for variable selection and interpretation. *J. Chemom.* **29**, 528–536. <https://doi.org/10.1002/cem.2736>.
28. Chong, J. et al. MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. *Nucleic Acids Res.* **46**, W486–W494 (2018). <https://doi.org/10.1093/nar/gky310>.
29. Barker, M. & Rayens, W. (2003). Partial least squares for discrimination. *J. Chemom.* **17**, 166–173. <https://doi.org/10.1002/cem.785>.

30. Davidsson, P. R., Kariola, T., Niemi, O. & Palva, E. T. Pathogenicity of and plant immunity to soft rot pectobacteria. *Front. Plant Sci.* **4**, 191 (2013). <https://doi.org/10.3389/fpls.2013.00191>.
31. Laurila, J. et al. Characterization of *Dickeya* strains isolated from potato and river water samples in Finland. *Eur. J. Plant Pathol.* **122**, 213–225 (2008). <https://doi.org/10.1007/s10658-008-9274-5>.
32. Bhat, K. A., Bhat, N. A., Mohiddin, F. A., Sheikh, P. A. & Wani, A. H. Studies on pectinase activities of isolates of *Erwinia carotovora* and *Rhizopus* sp. causing soft rot in cabbage (*Brassica oleracea* var. *capitata* L.). *Afr. J. Agric. Res.* **7**, 6062–6067 (2012). <https://doi.org/10.5897/AJAR12.1215>.
33. Gerhardt, N., Birkenmeier, M., Sanders, D., Rohn, S. & Weller, P. Resolution-optimized headspace gas chromatography-ion mobility spectrometry (HS-GC-IMS) for non-targeted olive oil profiling. *Anal. Bioanal. Chem.* **409**, 3933–3942 (2017). <https://doi.org/10.1007/s00216-017-0338-2>.
34. Tabachnick, B. G., Fidell, L. S. & Ullman, J. B. *Using multivariate statistics*, 5. (Pearson, 2007).
35. Thenkabail, P. S. et al. Selection of hyperspectral narrowbands (HNBS) and composition of hyperspectral twoband vegetation indices (HVIs) for biophysical characterization and discrimination of crop types using field reflectance and Hyperion/EO-1 data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **6**, 427–439 (2013). <https://doi.org/10.1109/JSTARS.2013.2252601>.
36. Andro, T. et al. Mutants of *Erwinia chrysanthemi* defective in secretion of pectinase and cellulase. *J. Bacteriol.* **160**, 1199–1203 (1984). <https://doi.org/10.1128/jb.160.3.1199-1203.1984>.
37. Murata, H., McEvoy, J. L., Chatterjee, A., Collmer, A. & Chatterjee, A. K. Molecular cloning of an *aepA* gene that activates production of extracellular pectolytic, cellulolytic, and proteolytic enzymes in *Erwinia carotovora* subsp. *carotovora*. *Mol. Plant–Microbe Interact.* **4**, 239–246 (1991). <https://doi.org/10.1094/MPMI-4-239>.
38. Esquerre, C., Gowen, A. A., Downey, G. & O'Donnell, C. P. Wavelength selection for development of a near infrared imaging system for early detection of bruise damage in mushrooms (*Agaricus bisporus*). *J. Near Infrared Spectrosc.* **20**, 537–546 (2012). <https://doi.org/10.1255/jnirs.1014>.
39. Nguyen-Do-Trong, N., Dusabumuremyi, J. C. & Saeys, W. Cross-polarized VNIR hyperspectral reflectance imaging for non-destructive quality evaluation of dried banana slices, drying process monitoring and control. *J. Food Eng.* **238**, 85–94 (2018). <https://doi.org/10.1016/j.jfoodeng.2018.06.013>.
40. Thenkabail, P. S., Gumma, M. K., Teluguntla, P., & Mohammed, I. A. Hyperspectral remote sensing of vegetation and agricultural crops. *Photogramm. Eng. Remote Sens.* **80**, 697–723 (2014).
41. Varmuza, K. & Filzmoser, P., *Introduction to multivariate statistical analysis in chemometrics*. (CRC Press, 2009).
42. Ambrose, A., Kandpal, L. M., Kim, M. S., Lee, W. H. & Cho, B. K. High speed measurement of corn seed viability using hyperspectral imaging. *Infrared Phys. Technol.* **75**, 173–179 (2016).
43. Williams, P. J., Geladi, P., Fox, G. & Manley, M. Maize kernel hardness classification by near infrared (NIR) hyperspectral imaging and multivariate data analysis. *Anal. Chim. Acta* **653**, 121–130 (2009). <https://doi.org/10.1016/j.aca.2009.09.005>.
44. Fraiwan, L., Lweesy, K., Khasawneh, N., Wenz, H. & Dickhaus, H. Automated sleep stage identification system based on time-frequency analysis of a single EEG channel and random forest classifier. *Comput. Methods Programs Biomed.* **108**, 10–19 (2012). <https://doi.org/10.1016/j.cmpb.2011.11.005>.
45. Amaratunga, D., Cabrera, J. & Lee, Y. S. Enriched random forests. *Bioinformatics* **24**, 2010–2014 (2008). <https://doi.org/10.1093/bioinformatics/btn356>
46. Keresztes, J. C. et al. Glare based apple sorting and iterative algorithm for bruise region detection using shortwave infrared hyperspectral imaging. *Postharvest Biol. Technol.* **130**, 103–115 (2017). <https://doi.org/10.1016/j.postharvbio.2017.04.005>.

47. Meyer, D., Leisch, F. & Hornik, K. The support vector machine under test. *Neurocomputing*. **55**, 169–186 (2003). [https://doi.org/10.1016/S0925-2312\(03\)00431-4](https://doi.org/10.1016/S0925-2312(03)00431-4).
48. Maldonado, S., Weber, R. & Basak, J. Simultaneous feature selection and classification using kernel-penalized support vector machines. *Inf. Sci.* **181**, 115–128 (2011). <https://doi.org/10.1016/j.ins.2010.08.047>.
49. Zhang, C. et al. Hyperspectral imaging analysis for ripeness evaluation of strawberry with support vector machine. *J. Food Eng.* **179**, 11–18 (2016). <https://doi.org/10.1016/j.jfoodeng.2016.01.002>.
50. Huang, M., Zhu, Q., Wang, B. & Lu, R. Analysis of hyperspectral scattering images using locally linear embedding algorithm for apple mealiness classification. *Comput. Electron. Agric.* **89**, 175–181 (2012). <https://doi.org/10.1016/j.compag.2012.09.003>.

Figures

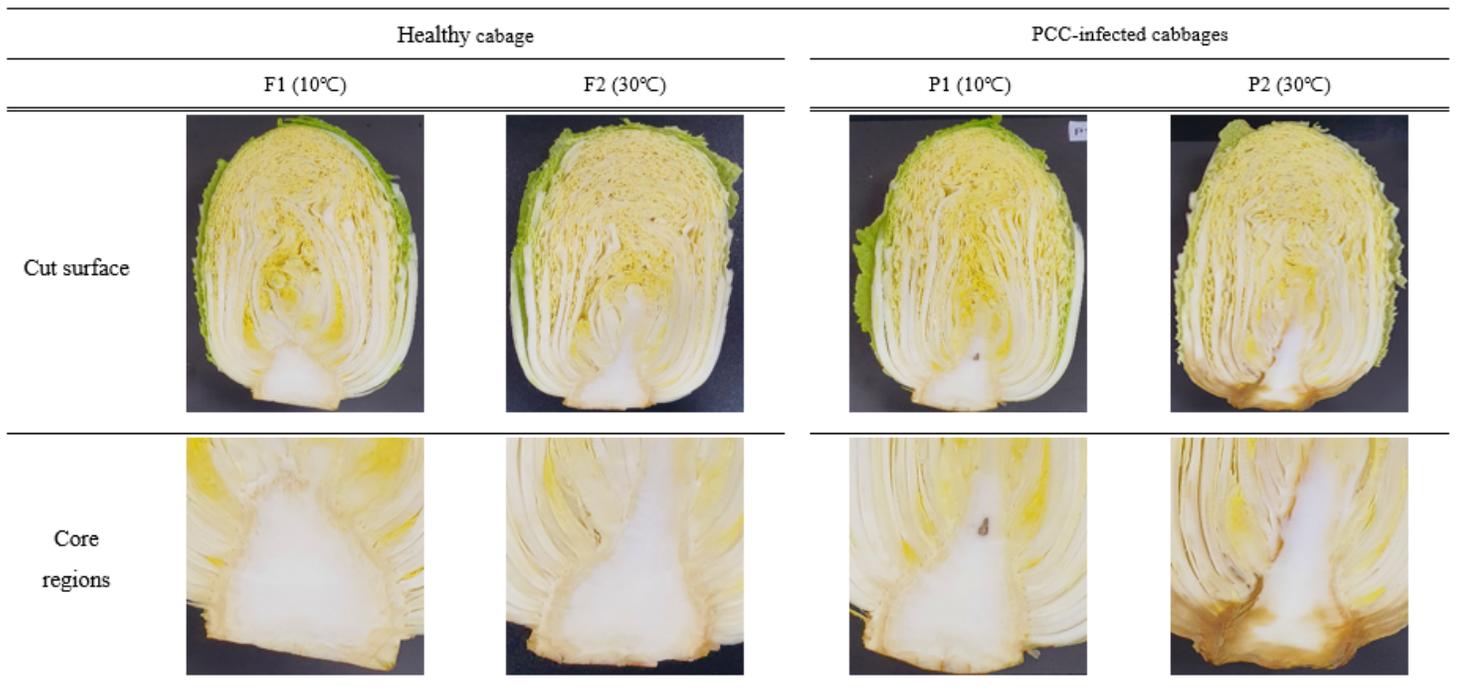


Figure 1

Comparison of *Pectobacterium carotovorum* subsp. *carotovorum* (PCC) bacterial counts and soft rot development between healthy cabbages and PCC-inoculated cabbages under different storage temperatures after 3 days.

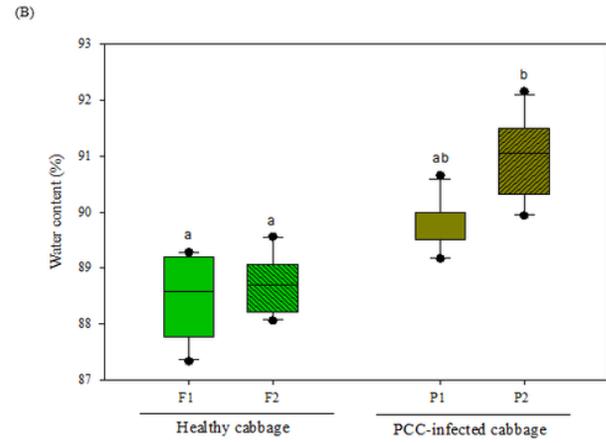
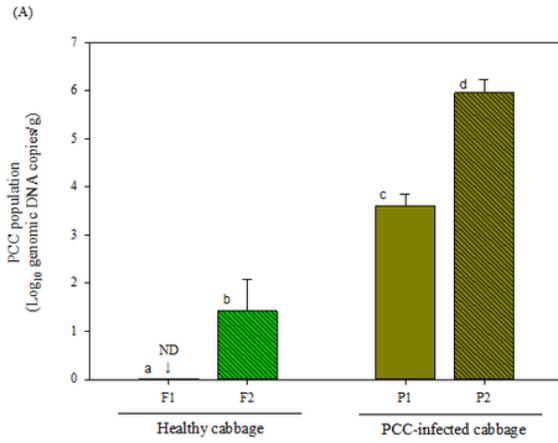


Figure 2

The quantity of *Pectobacterium carotovorum* ssp. *carotovorum* cells recovered from napa cabbage (A) and water content of cabbage samples (B). Healthy napa cabbage stored at 5 °C (F1), healthy napa cabbage stored at 30 °C (F2), napa cabbage inoculated with PCC and stored at 5 °C (P1), and napa cabbage inoculated with PCC and stored at 30 °C (P2).

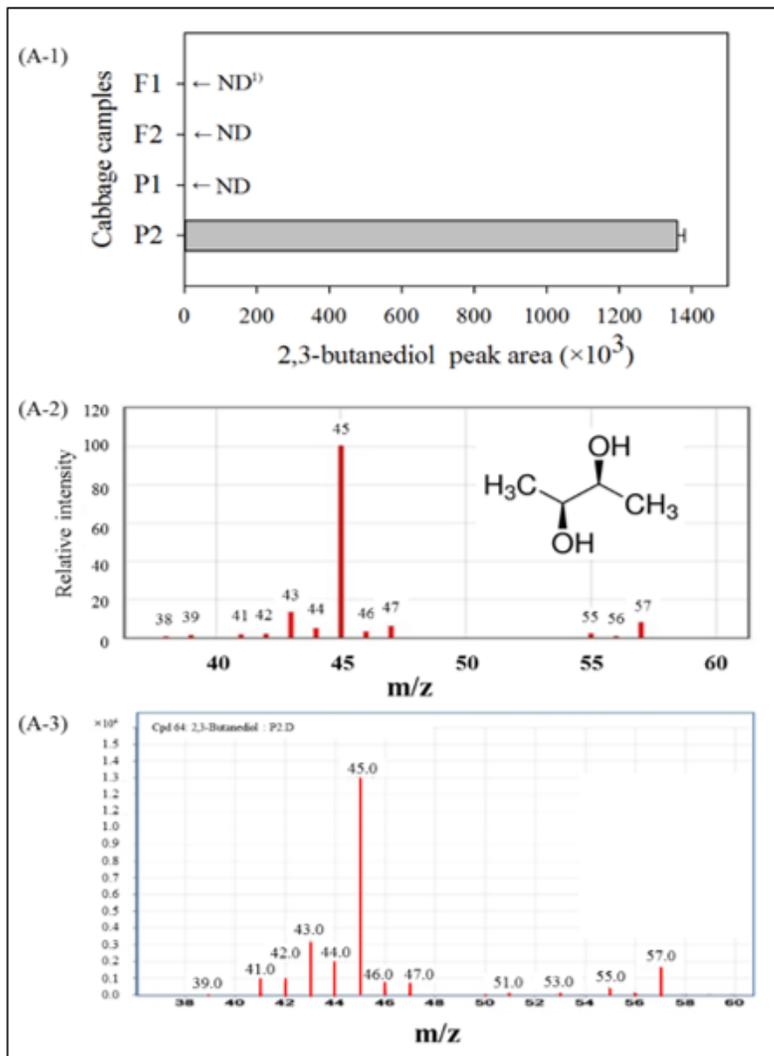


Figure 3

A headspace solid-phase microextraction coupled with gas chromatography–mass spectrometry (HS-SPME-GC-MS) chromatogram. (A-1) Volatile compounds in four kinds of napa cabbage samples at different conditions. The area values were obtained from values integrated from area of peaks on Total Ion Chromatogram Total Ion Chromatogram. Healthy napa cabbage stored at 5 °C (F1), healthy napa cabbage stored at 30 °C (F2), napa cabbage inoculated with PCC and stored at 5 °C (P1), and napa cabbage inoculated with PCC and stored 30 °C (P2); (A-2) A standard solution containing 2,3-butanediol; (A-3) 2,3-Butanediol in napa cabbage samples showing soft rot disease symptom.

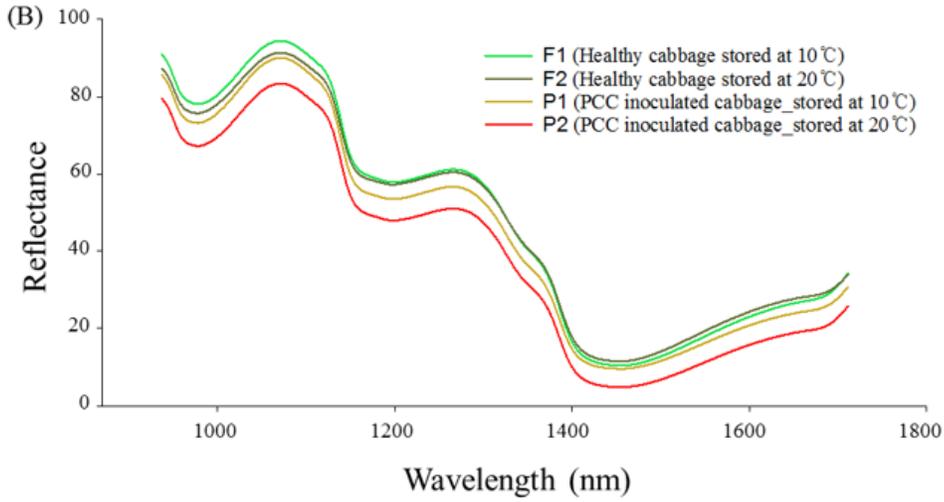
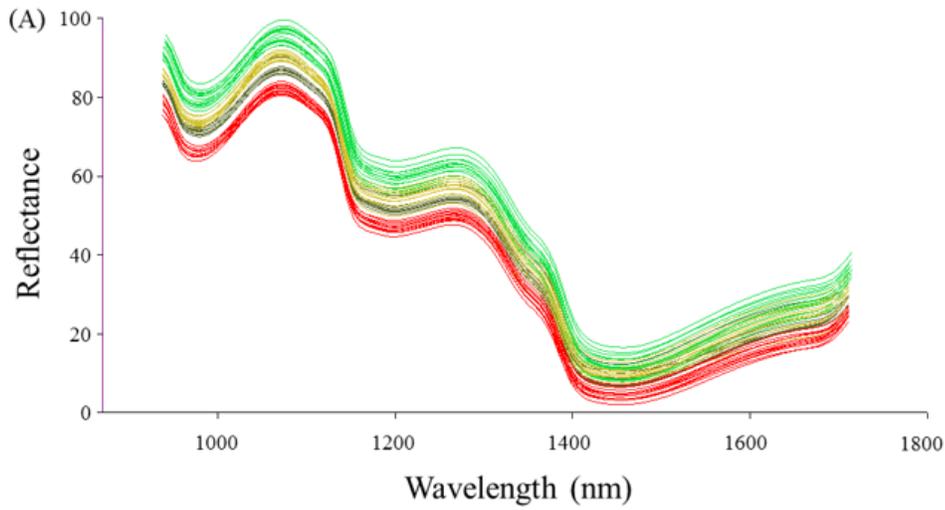


Figure 4

The reflectance spectra of napa cabbages of all samples (A) and average spectra for different napa cabbage sample groups (F1, F2, P1, and P2) (B).

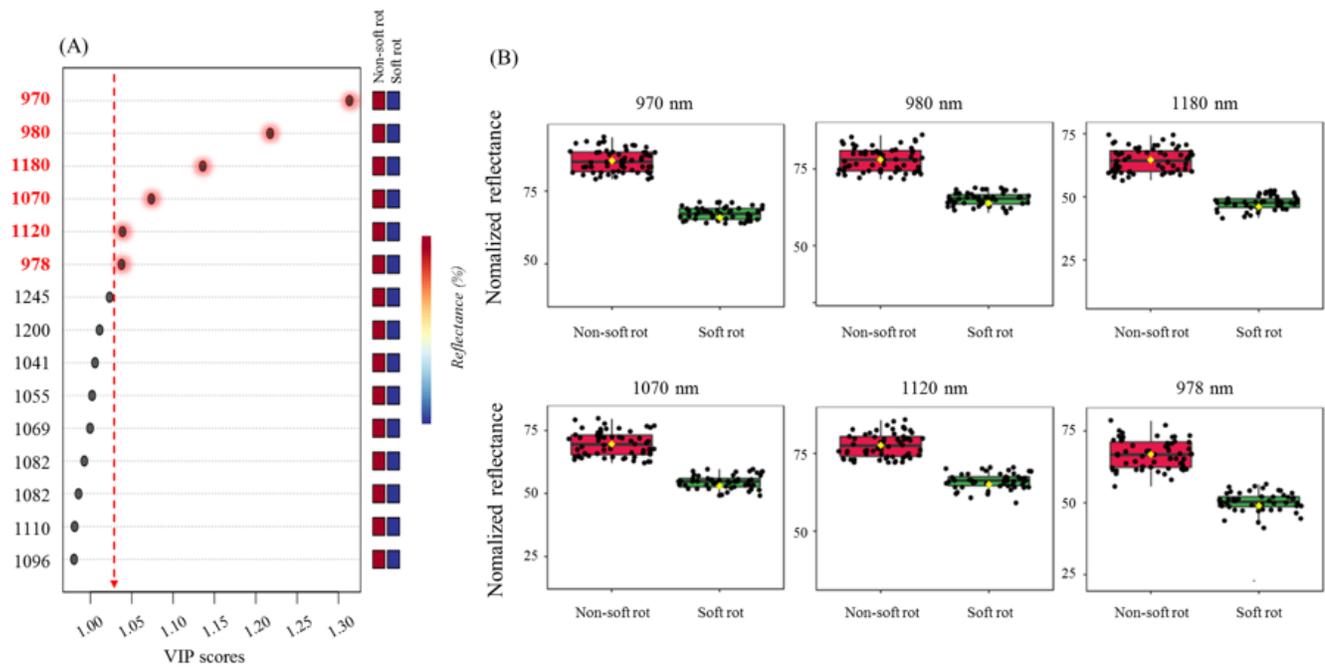


Figure 5

Significant wavelengths by PLS-DA model that efficiently classify soft rot disease infection (A) and the comparison of boxplot of the top six significant relatively important wavelengths ($p < 0.05$) in four groups of napa cabbage samples (B).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementaryfile.docx](#)